TOTAL BUDGET: DISC 3.1, 2016	
Funded	$2,000,000
Not Funded	$8,736,760

App #	Title	Score	Median	SD	Low	High	Budget	Tier	T1	T2
DISC3.1-09167	Genomic characterization of the CIRM hu	88	90	6	75	95	$2,000,000	1	9	2
DISC3.1-09024	Comprehensive Genetic Profiling of hiPS(	85	85	7	75	100	$1,999,998	2	8	3
DISC3.1-09175	CIRM iPSC Biorepository: Elucidating Cau	81	80	5	75	95	$2,000,000	2	1	9
DISC3.1-09176	Genetic Profiling of CIRM hiPSC through (	74	75	6	60	80	$2,737,608	2	0	11
DISC3.1-09177	Genetic Profiling of CIRM's hiPSC Reposit	68	70	6	50	70	$1,999,154	2	0	10

DISCOVERY

Public Summary for DISC3.1-09024

Application #	DISC3.1-09024
Title (as written by the applicant)	Comprehensive Genetic Profiling of hiPSCs
Major Proposed Activities (as written by the applicant)	• Comprehensive Genetic Analysis of HumanCore-24 BeadChip Data from 3000 hiPSC. • Selection and Whole Genome Sequencing of 1000 samples. • Analysis of Whole Genome Sequencing Data to Characterize the Types and Prevalence of Somatic Mutations Present in hiPSC Derived from Different Source Tissues.
Statement of Benefit to California (as written by the applicant)	The 11 diseases represented in this hiPSC collection were chosen based on their broad impact in California. Providing genetic data to supplement this well described collection will add value to the ability to select appropriate lines in order to model these genetically complex diseases. Analysis of somatic mutations that arise during reprogramming will inform future reprogramming experiments, and help inform use of hiPSC in the lab and clinic.
Funds Requested	$1,999,998
GWG Recommendation	***Not recommended for funding***
CIRM Team Recommendation	***CIRM Team concurs with the GWG's recommendation.***

DISCOVERY

Scoring Data

Final Score: 85

Up to 15 scientific members of the GWG score each application. The final score for an application is the average of the individual member scores. Additional parameters related to the score are shown below. A score of 85-100 indicates that an application is “Recommended for funding” and a score of 1-84 indicates that an application is “Not recommended for funding.” For programs that can fund only one application, such as this one, only the application with the highest mean score within the “Recommended for funding” category will be recommended.

Median	85
Standard Deviation	7
Highest	100
Lowest	75
Count	11
Number of reviewers who scored 85-100	8
Number of reviewers who scored 1-84	3

Score Influences

Proposals were evaluated and scored based on the criteria shown below, which are also described in the RFA. The scientific members of the GWG were asked to indicate how their evaluation of the proposal against each criterion influenced their overall score. The total number of reviewers indicating a positive, negative, or neutral influence for each criterion is shown.

Criterion	Positive Influence	Negative Influence	Neutral Influence
Does the proposal have a potential for impact?	6	1	4
Is the rationale sound?	5	2	4
Is the proposal well planned and designed?	4	2	5
Is the proposal feasible?	8	0	3

Reviewer Comments

The following is a compilation of comments provided by multiple reviewers following the panel's discussion and scoring of the application. All reviewers were asked to provide brief bullets on key strengths, concerns, or recommendations related to the proposal that CIRM compiled and edited for clarity.

Strengths

- The Principal Investigator (PI) is competent and runs a large core laboratory. The PI has extensive experience and an excellent track record running projects of this type.
- The mechanisms are in place to accomplish the goals. The PI is strong with demonstrated scientific excellence.
- The team will use SNP data to select 1000 samples and perform whole genome sequencing (WGS). This is a strong team with good expertise. The integration with many other existing WGS data sets is a strength.
- Selected samples that include the origin and iPSC generated cells (ideally at different passages) would be highly informative and would provide an innovative and truly unique resource.
- The partnership between the PI and a company is a strength.

Concerns

- There are mild concerns about the proposed method of choosing the subset of the lines to sequence. Disease balance and general genetic diversity should be the criteria instead of an effort to take into account as many slight risk factors (e.g., GWAS alleles) as possible.
- A complex approach to evaluating somatic mutations is presented, and the PI spends a page on this. The details of enrichment for candidate mutations are critical but are not provided.
- The statistical power of this method is unclear. How the team will differentiate mutations from technical error is unclear. Typically, one need for such a project is validation of candidate mutations, which is expensive and has no budget for such work in the application.
- The normal cell lines from a collaborator will not necessarily be collected and sequenced with the same technology and software versions at the same time, so a comparison may be meaningless.
- The plan for distribution to the public is weak.
- In the absence of donor data (i.e. iPSC origin) the sequencing data of iPSC cells are of very limited value.

Additional Comments

- The paragraph on "Enrichment analysis of aberrant methylation sites in hiPSC" seems extraneous.

DISCOVERY

Public Summary for DISC3.1-09167

Application #	DISC3.1-09167
Title (as written by the applicant)	Genomic characterization of the CIRM human induced pluripotent stem cell (hiPSC) repository
Major Proposed Activities (as written by the applicant)	• Sample Selection: Disease, population features, genotype and availability information will be used to select a diverse set of 750-1000 lines for comprehensive analysis by whole genome sequencing. • Library preparation: Established protocols for Illumina HiSeq or 10X Genomics will be used by our commercial sequencing partner, Macrogen, to prepare whole genome sequencing libraries. • Sequencing: Samples will be sequenced on the Illumina HiSeq X Ten platform, the current benchmark in whole genome sequencing, to generate coverage of the entire genome at an average depth of 30X. • Variant calling: Sequence analysis will take place on the cloud, using several methods. Results will include DNA substitutions, deletions, insertions, genome rearrangements and haplotype information. • Data Sharing: Files summarizing all variants identified for each sample will be sent to DbGaP, Coriell and the CIRM Data Coordination and Management site. Raw data will be available through DbGaP.
Statement of Benefit to California (as written by the applicant)	This iPSC bank can generate medical breakthroughs, advance California's biotech and pharmaceutical industries and lead to healthier Californians, but only if we understand their genetic features. Which genetic aberrations underlie disease phenotypes? Which arise during iPSC generation and culture? Which represent normal variation between people? Through whole genome analysis we strive to understand the relationship between specific variants and the etiology of the diseases these lines represent.
Funds Requested	$2,000,000
GWG Recommendation	***Recommended for funding***
CIRM Team Recommendation	***CIRM Team concurs with the GWG's recommendation.***

Scoring Data

Final Score: 88

Up to 15 scientific members of the GWG score each application. The final score for an application is the average of the individual member scores. Additional parameters related to the score are shown below. A score of 85-100 indicates that an application is "Recommended for funding" and a score of 1-84 indicates that an application is "Not recommended for funding." For programs that can fund only one application, such as this one, only the application with the highest mean score within the "Recommended for funding" category will be recommended.

Median	90
Standard Deviation	6
Highest	95
Lowest	75
Count	11
Number of reviewers who scored 85-100	9
Number of reviewers who scored 1-84	2

Score Influences

Proposals were evaluated and scored based on the criteria shown below, which are also described in the RFA. The scientific members of the GWG were asked to indicate how their evaluation of the proposal against each criterion influenced their overall score. The total number of reviewers indicating a positive, negative, or neutral influence for each criterion is shown.

Criterion	Positive Influence	Negative Influence	Neutral Influence
Does the proposal have a potential for impact?	7	1	3
Is the rationale sound?	6	1	4
Is the proposal well planned and designed?	8	0	3
Is the proposal feasible?	8	0	3

Reviewer Comments

The following is a compilation of comments provided by multiple reviewers following the panel's discussion and scoring of the application. All reviewers were asked to provide brief bullets on key strengths, concerns, or recommendations related to the proposal that CIRM compiled and edited for clarity.

Strengths

- The haplotype data could be interesting, e.g., for studying compound heterozygosity.
- Outsourcing data generation is likely to generate high efficiency and uniform unbiased data.
- There is a great plan for disseminating data to the public, both for high-throughput users and for casual users.
- Selected samples that include the origin and iPSC generated cells (ideally at different passages) would be highly informative and would provide an innovative and truly unique resource.
- This is a strong proposal. The phased genomes add value. It is, however, unclear how many will be provided, e.g., there are large ranges 750-1000 whole genome sequences and 40-750 of what is being delivered.
- Whole genome sequencing approach is a strength
- Strong computational and data handling in place

Concerns

- Although this is a strong proposal, there are some reservations about whether time and effort commitments are adequate. The team may need more FTE support for personnel.
- In the absence of donor data (i.e., iPSC origin) the sequencing data of iPSC cells are of very limited value.

DISCOVERY

Public Summary for DISC3.1-09175

Application #	DISC3.1-09175
Title (as written by the applicant)	CIRM iPSC Biorepository: Elucidating Causal Genetic Variants by Whole Exome Sequencing
Major Proposed Activities (as written by the applicant)	• Obtain DNA from each sample and assess DNA quality. • Construct DNA libraries. • Read the DNA sequence. • Analyze the data. • Share the data with the scientific community.
Statement of Benefit to California (as written by the applicant)	Our work will identify the underlying genetic causes of complex diseases, may lead to the development of novel treatments that can prevent the disease from happening or progressing. Improving the outcome of the health disease and providing new treatments will go a long way to helping a large group of Californians to lead healthier lives. Also, the outcome of our research could potentially improve the California health care system by reducing the cost of the health care burden.
Funds Requested	$2,000,000
GWG Recommendation	***Not recommended for funding***
CIRM Team Recommendation	***CIRM Team concurs with the GWG's recommendation.***

Scoring Data

Final Score: 81

Up to 15 scientific members of the GWG score each application. The final score for an application is the average of the individual member scores. Additional parameters related to the score are shown below. A score of 85-100 indicates that an application is “Recommended for funding” and a score of 1-84 indicates that an application is “Not recommended for funding.” For programs that can fund only one application, such as this one, only the application with the highest mean score within the “Recommended for funding” category will be recommended.

Median	80
Standard Deviation	5
Highest	95
Lowest	75
Count	10
Number of reviewers who scored 85-100	1
Number of reviewers who scored 1-84	9

Score Influences

Proposals were evaluated and scored based on the criteria shown below, which are also described in the RFA. The scientific members of the GWG were asked to indicate how their evaluation of the proposal against each criterion influenced their overall score. The total number of reviewers indicating a positive, negative, or neutral influence for each criterion is shown.

Criterion	Positive Influence	Negative Influence	Neutral Influence
Does the proposal have a potential for impact?	6	2	2
Is the rationale sound?	4	3	3
Is the proposal well planned and designed?	5	1	4
Is the proposal feasible?	8	0	2

DISCOVERY

Reviewer Comments

The following is a compilation of comments provided by multiple reviewers following the panel's discussion and scoring of the application. All reviewers were asked to provide brief bullets on key strengths, concerns, or recommendations related to the proposal that CIRM compiled and edited for clarity.

Strengths

- The proposal to use exome sequencing on 3000 samples makes this a highly significant application.

Concerns

- Some reviewers preferred performing whole-genome sequencing (WGS) on a smaller number of lines rather than the proposed exomes.
- Some concern was expressed about the lack of documented experience of this group in producing and analyzing data on the scale required.
- The quality of exome data is highly variable. For example, the coverage near the ends of capture regions can be variable, leading to complex computational and statistical considerations in interpreting the absence of calls, particularly when comparing exome data to other data types or other types of exome data.
- There was no discussion of batch effects, e.g., how to prevent, detect, and mitigate them.
- While comprehensive (all 3000 samples), the exome approach is a limitation, notwithstanding the applicant's argument. The budget appears to be insufficient and would likely require extra sources beyond CIRM funding.
- Whole-exome sequencing on all 3000 lines may uncover disease mutations but studies will be inadequately powered to draw genotype-phenotype conclusions.
- The proposed approach will miss noncoding regions that may be crucial to disease process.

Additional Comments

- Many of the needs for analyzing iPSC lines can only be achieved with WGS, e.g., regulation, genome rearrangements, mutation rates, etc.
- It is much harder to integrate and compare exome data from different labs. It can be hard for WGS, but much harder for exome. Coverage profiles at particular positions can vary highly, for example, due to different capture technologies or even versions of the same technology. Much of the value of the CIRM data will be gained from comparisons with public data.

DISCOVERY

Public Summary for DISC3.1-09176

Application #	DISC3.1-09176
Title (as written by the applicant)	Genetic Profiling of CIRM hiPSC through Genome-Wide Sequencing and Comparison to Multiple Control Datasets
Major Proposed Activities (as written by the applicant)	• Array genotyping data rare and common variant analysis (Year 1). • Generate exome data for 1st 1000 hiPSC samples (Year 1). • Integrate array and exome genotype data (Years 1 and 2). • Integrate array and exome genotype data (Year 2). • Map exome data to reference human genome, call variants, re-call variants as cohort grows (Years 1 and 2). • Deposit exome data to public repositories (end of Year 1, mid-Year 2, end of Year 2).
Statement of Benefit to California (as written by the applicant)	Understanding the connection between genome variation and the molecular mechanisms underlying disease is critical to the development of new therapies for disease and molecular markers to measure patient improvement or progression. Reading genome variants provides a window into the molecular potential of a patient's cells. For the CIRM hiPSC, we will generate a fine-grained exome map of variation to complement array-based maps. Comparision to large control cohorts will highlight disease genes.
Funds Requested	$2,737,608
GWG Recommendation	***Not recommended for funding***
CIRM Team Recommendation	***CIRM Team concurs with the GWG's recommendation.***

DISCOVERY

Scoring Data

Final Score: 74

Up to 15 scientific members of the GWG score each application. The final score for an application is the average of the individual member scores. Additional parameters related to the score are shown below. A score of 85-100 indicates that an application is “Recommended for funding” and a score of 1-84 indicates that an application is “Not recommended for funding.” For programs that can fund only one application, such as this one, only the application with the highest mean score within the “Recommended for funding” category will be recommended.

Median	75
Standard Deviation	6
Highest	80
Lowest	60
Count	11
Number of reviewers who scored 85-100	0
Number of reviewers who scored 1-84	11

Score Influences

Proposals were evaluated and scored based on the criteria shown below, which are also described in the RFA. The scientific members of the GWG were asked to indicate how their evaluation of the proposal against each criterion influenced their overall score. The total number of reviewers indicating a positive, negative, or neutral influence for each criterion is shown.

Criterion	Positive Influence	Negative Influence	Neutral Influence
Does the proposal have a potential for impact?	5	3	3
Is the rationale sound?	4	3	4
Is the proposal well planned and designed?	0	7	4
Is the proposal feasible?	4	2	5

DISCOVERY

Reviewer Comments

The following is a compilation of comments provided by multiple reviewers following the panel's discussion and scoring of the application. All reviewers were asked to provide brief bullets on key strengths, concerns, or recommendations related to the proposal that CIRM compiled and edited for clarity.

Strengths

- Strong PI in the computational biology field with the scientific skills to oversee data analysis and potentially add creative touches to the product.

Concerns

- Underdeveloped as an actual project--dependent on making adequate arrangements with core laboratories to do the work and to hire new personnel to do the analyses.
- This project has a single critical (and lethal) flaw: it proposes to do sequencing at multiple sites, using different machines. This has the potential to introduce batch effects.
- Batch effects are the single biggest danger to a project to sequence thousands of genomes. It is not necessary to run this risk.
- New exome data for up to 2000 samples will be integrated with existing array genotype data for 3000 samples. Concerns are the limited data acquired through the exomes. While powerful, a concern is the many non-coding regions are being missed.
- The cost for each exome listed in the proposal is not competitive.

Additional Comments

- There is a discussion of Primary Open-Angle Glaucoma (POAG), but it isn't sufficiently addressed. It could be helpful to harness a particular computational focus to this sequencing project, as it adds an interaction between end-users and data-generation.
- This is (or could be) a strength of this project. But the details seem to have been forgotten after the introduction. Perhaps because, in practical terms, there is not much budget for this sort of activity.

DISCOVERY

Public Summary for DISC3.1-09177

Application #	DISC3.1-09177
Title (as written by the applicant)	Genetic Profiling of CIRM's hiPSC Repository
Major Proposed Activities (as written by the applicant)	• Obtain hIPSC lines from the CIRM biobank and prepare DNA from them for sequencing. • Analyze the existing data on single nucleotide polymorphisms (SNPs) in the iPSC lines to learn about the genomic stability of the cells and an overview of their genetic risk factors. • Use bioinformatics tools to select a subset of individual's iPSCs for whole genome sequencing. Priority will be given to underrepresented minority families. Sequence DNA. • Submit whole genome sequencing data to the NIH's database of Genotypes and Phenotypes (dbGaP) to archive and distribute the data relevant to interaction of genotype and disease. • Use established best bioinformatic practices to analyze sequence data for nucleotide variants associated with disease risk and cancer. • Work with the biobank and the stem cell research community to provide the most useful genomic data on each iPSC line on the biobank's website.
Statement of Benefit to California (as written by the applicant)	California has one of the most diverse populations in the US. Linking clinical data with genomic data about the iPSCs in CIRM's collection will provide a tremendous resource for research to understand human disease as well as a means to identify and test novel therapies. Californians will benefit from the unprecedented wealth of knowledge, which will lead to improvements in the health of its citizens.
Funds Requested	$1,999,154
GWG Recommendation	***Not recommended for funding***
CIRM Team Recommendation	***CIRM Team concurs with the GWG's recommendation.***

Scoring Data

Final Score: 68

Up to 15 scientific members of the GWG score each application. The final score for an application is the average of the individual member scores. Additional parameters related to the score are shown below. A score of 85-100 indicates that an application is “Recommended for funding” and a score of 1-84 indicates that an application is “Not recommended for funding.” For programs that can fund only one application, such as this one, only the application with the highest mean score within the “Recommended for funding” category will be recommended.

Median	70
Standard Deviation	6
Highest	70
Lowest	50
Count	10
Number of reviewers who scored 85-100	0
Number of reviewers who scored 1-84	10

Score Influences

Proposals were evaluated and scored based on the criteria shown below, which are also described in the RFA. The scientific members of the GWG were asked to indicate how their evaluation of the proposal against each criterion influenced their overall score. The total number of reviewers indicating a positive, negative, or neutral influence for each criterion is shown.

Criterion	Positive Influence	Negative Influence	Neutral Influence
Does the proposal have a potential for impact?	5	2	3
Is the rationale sound?	3	4	3
Is the proposal well planned and designed?	0	4	6
Is the proposal feasible?	5	1	4

DISCOVERY

Reviewer Comments

The following is a compilation of comments provided by multiple reviewers following the panel's discussion and scoring of the application. All reviewers were asked to provide brief bullets on key strengths, concerns, or recommendations related to the proposal that CIRM compiled and edited for clarity.

Strengths

- This is a good proposal by competent investigator.
- The creation of a web portal is nice.
- Whole genome sequencing will enable capture of noncoding sequences important in disease.

Concerns

- It is not fully competitive with other proposals for the amount of data that would be produced for the available budget.
- There are a low number of samples to be sequenced compared to total repository.
- The focus on DNA quality is good, but expensive.
- The decision to grow cells in lab rather than use DNA available not well-justified and is resource-intensive.
- This is a straightforward proposal but lacking in key management issues.

Additional Comments

- The goal of the proposed work is to perform and present an in depth genetic characterization of the CIRM hiPSC line collection and integrate it with the available clinical data. Industrial scale sequencing resources provided by a subcontracted company to cost efficiently analyze ~500 whole genome sequences from the CIRM iPSC collection. Although the team is strong, the limited sample number and high costs of adding the cell culture aspect reduce the power of this application.